Université de Versailles
http://dumas.perso.math.cnrs.fr/MINT-optim.html

Année 2019-2020
semestre 1

CC2: optimisation numérique, correction

Exercice 1

On s'intéresse à la minimisation d'une fonction quadratique $J : \mathbb{R}^n \to \mathbb{R}$ telle que

$$J(x) = \frac{1}{2}\,{}^t x A x - {}^t b x$$

avec A une matrice symétrique définie positive de taille n et $b \in \mathbb{R}^n$, sous la contrainte

$$Cx = 0$$

avec C une matrice de taille $p \times n$ ($p < n$), de rang maximal p.

1. Montrer que J admet un unique minimum global x^* tel que $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^p$ est solution unique de

$$\begin{cases} Ax^* + {}^t C\lambda^* = b \\ Cx^* = 0 \end{cases}$$

(3 pts) On montre tout d'abord que J est coercive et strictement convexe sur l'ensemble admissible Ω (vu en TD). Celui-ci est par ailleurs convexe et fermé. La coercivité de J et le caractère fermé de Ω montre l'existence du minimum et la stricte convexité de J et le caractère convexe de Ω son unicité. Le point (x^*, λ^*) vérifie en particulier les conditions de KKT qui s'écrivent comme indiqué car

$$\mathcal{L}(x, \lambda) = J(x) +^t (Cx)\lambda$$

2. On propose d'approcher x^* par la méthode suivante pour$(x_0, \lambda_0) \in \mathbb{R}^n \times \mathbb{R}^p$ donnés :

$$\begin{cases} x_{k+1} = x_k - \rho_1(Ax_k - b +^t C\lambda_k) \\ \lambda_{k+1} = \lambda_k + \rho_1\rho_2 C x_{k+1} \end{cases}$$

Jusitifier l'introduction de cette méthode à l'aide du Lagrangien du problème.

(1 pt) Il s'agit d'une méthode inspirée d'Uzawa qui consiste à recherche le point selle de $\mathcal{L}$ en effectuant successivement une itération de gradient (desecendant) en x :

$$x_{k+1} = x_k - \rho_1 \nabla_x \mathcal{L}(x_k, \lambda_k)$$

et une itération de gradient ascendante en λ :

$$\lambda_{k+1} = \lambda_k + \rho_1\rho_2 \nabla_\lambda \mathcal{L}(x_{k+1}, \lambda_k)$$

3. On note $\beta = |||I - \rho_1 A|||$ la norme subordonnée de $I - \rho_1 A$. On rappelle que si $x \in \mathbb{R}^n$,

$$||Ax|| \leq |||A|||.||x||$$

et que si A est symétrique définie positive, alors $|||A|||$ est égal à sa plus grande valeur propre.

4. Montrer que si ρ_1 est suffisamment petit, alors $\beta < 1$.
   (1 pt.) On note 0<$\mu_1 \leq ... \leq \mu_n$ les valeurs propres de A. Lorsque ρ_1 est suffisamment petit, $I - \rho_1 A$ est définie positive et sa plus grande valeur propre est $1 - \rho_1\mu_1 \in]0,1[$. On en déduit que $\beta = |||I - \rho_1 A||| < 1$
5. Montrer que
   $$||\lambda_{k+1}-\lambda^*||^2 \leq ||\lambda_k-\lambda^*||^2+(\rho_1\rho_2)^2|||^tCC|||.||x_{k+1}-x^*||^2+2\rho_1\rho_2 <^t C(\lambda_k-\lambda^*), x_{k+1}-x^* >$$
   (2 pts.) On écrit
   $$\lambda_{k+1}-\lambda^* = \lambda_{k+1}-\lambda_k+\lambda_k-\lambda^* = \rho_1\rho_2 C x_{k+1}+(\lambda_k-\lambda^*) = \rho_1\rho_2 C(x_{k+1}-x^*)+(\lambda_k-\lambda^*)$$
   puis on prend la norme au carré et on développe avec Pythagore :
   $$||\lambda_{k+1}-\lambda^*||^2 = ||\lambda_k-\lambda^*||^2 + 2 < \rho_1\rho_2 C(x_{k+1}-x^*), \lambda_k-\lambda^*) + ||\rho_1\rho_2 C(x_{k+1}-x^*)||^2$$
   Le dernier terme se récrit :
   $$< C(x_{k+1}-x^*), C(x_{k+1}-x^*) > = <^t CC(x_{k+1}-x^*), (x_{k+1}-x^*) >$$
   On conclut avec Cauchy Schwarz et la définition d'une triple norme.
6. On admet alors que si ρ_2 est suffisamment petit, il existe $\gamma > 0$ tel que
   $$\gamma||x_{k+1}-x^*||^2 \leq \left(\frac{||\lambda_k-\lambda^*||^2}{\rho_2}+\beta||x_k-x^*||^2\right) - \left(\frac{||\lambda_{k+1}-\lambda^*||^2}{\rho_2}+\beta||x_{k+1}-x^*||^2\right)$$
   Montrer que pour un tel choix de paramètres, on a $\lim_{k\to+\infty} x_k = x^*$.
   (2 pts.) On somme les relations de 0 à N et on obtient :
   $$\sum_{k=0}^{N}\gamma||x_{k+1}-x^*||^2 \leq \frac{||\lambda_0-\lambda^*||^2}{\rho_2}+\beta||x_0-x^*||^2 - 0$$
   On se trouve en présence d'une série à termes positifs majorée. Elle est ainsi convergente et son terme général tend donc vers 0 ce qui permet de conclure.
7. Que peut-on dire de la suite $(\lambda_k)_{k\in\mathbb{N}}$?
   (2 pts.) En reprenant la relation définissant x_{k+1}, on a ${}^tC\lambda_k$ tend vers 0. En composant par C, on en déduit que $C^tC\lambda_k$ tend vers 0. Comme C^tC est une matrice carré inversible (car C est de rang maximal), on en déduit que λ_k tend vers 0.

Exercice 2

Soit la fonction.
$$J(x,y) = 3x^4 + y^2 - 3xy$$
à minimiser sur l'ensemble
$$D = \{(x,y) \in \mathbb{R}^2, 1 \leq x \quad \text{et} \quad y \geq 3\}$$

1. Montrer que ce problème admet une unique solution (qu'on ne cherchera pas à déterminer).
   (2 pts.) On remarque tout d'abord que l'ensemble admissible est convexe et fermé. Il reste à montrer que J est coercive et strictement convexe sur Ω. Pour la stricte convexité, on calcule le Hessien de J. Celui-ci est égal à $\begin{pmatrix} 36x^2 & -3 \\ -3 & 2 \end{pmatrix}$. Lorsque $x \geq 1$, cette matrice est bien définie positive par critère de Sylvester ($72x^2 - 9 \geq 0$). Pour la coercivité, pour $x \geq 1$, on minore J par la fonction quadratique $3x^2 + y^2 - 3xy$ qui est elle même coercive en calculant son Hessien également défini positif par Sylvester.

2. Ecrire, de manière explicite une itération de l'algorithme du gradient projeté (à pas constant) sur cet exemple.

   (2 pts.) On a

$$x_{k+1} = \max(1, x_k - \rho(12x_k^3 - 3y_k))$$

   et

$$y_{k+1} = \max(3, y_k - \rho(2y_k - 3x_k))$$

Exercice 3

Sous sa forme la plus élémentaire, une stratégie d'évolution s'écrit ainsi :

```
////////////////////////////////////////////////
n=10;lambda=50;
mu=5;
Ngen=200;
sigma=0.01;
/////////////////////////////////////////////////
function y=sphere(x)
    y=sum(x.*x);
endfunction
///////////////////////////////////////////////////
Ap=2*rand(mu,n+1)-ones(mu,n+1)// initialisation
Ae=zeros(lambda,n+1)
//
for i=1:Ngen
//
  X=1/mu*sum(Ap(:,1:n),'r')  // croisement
  for i=1:lambda
     Ae(i,1:n)=X+sigma*rand(1,n,'normal') // mutation
     Ae(i,n+1)=sphere(Ae(i,1:n))   // evaluation
  end
  [a,b]=gsort(-Ae(:,n+1))
  Ae=Ae(b,:)
  Ap=Ae(1:mu,:)  // selection virgule (enfants seulement)
end
```

1. Que représentent les paramètres `n`, `lambda` et `mu` ?

   (1 pt.) On a `n` *: nombre de variables,* `lambda` *: nombre d'enfants et* `mu` *: nombre de parents*

2. On souhaite enlever l'opérateur de croisement. A la place, les enfants sont issus d'un parent tiré (à chaque fois) aléatoirement. Comment proposez vous de modifier le code ?

   (2 pt.) `p=int(lambda*rand())+1, Ae(i,1:n)=Ap(1,1:n)+sigma*rand(1,n,'normal')`

3. Que valent `a` et `b` si on tape dans Scilab : `[a,b]=gsort([3.3,4,2,6,1])` ?

   (2 pts.) On a `a=[6,4,3.3,2,1]` *et* `b=[4,2,1,3,5]`

4. Quel sera le temps d'éxécution de ce code si on remplace la fonction `sphere` par une fonction nécessitant pour chaque évaluation 1 minute de calcul ?

   (2 pts.) Il faut effectuer $200 * 50 = 10000$ *appels à la fonction et donc il faut* 10000 *minutes*